

# Heuristic decomposition of cones in molecular conformational space.

Jacques Gabarro-Arpa

Ecole Normale Supérieure de Cachan, LBPA,CNRS UMR 8113  
61, Avenue du Président Wilson, 94235 Cachan cedex, France

Email: jga@lbpa.ens-cachan.fr

*Abstract* In a previous work [physics/061108v2], it was shown that the volume spanned by a molecular system in its conformational space can be effectively bounded by a polyhedral cone, this cone is described by means of a simple combinatorial formula. On the other hand it was constructed a transversal graph structure encoding the region of conformational space accessible to the system. From the information in this graph, it is possible to decompose the main cone into a hierarchy of smaller ones that are more manageable, and are progressively more tightly bound to the region in which the system evolves.

*Keywords* **Molecular Conformational Space, Polyhedral Cone, Poset, Molecular Dynamics**

*Mathematics Subject Classification:* 52B11, 52B40, 65Z05

*PACS:* 02.70.Ns

# I. PARTITION SEQUENCES

A partition of molecular conformational space (thereafter referred as  $CS$ ) was defined and studied in [1-4] based on the notion of **dominance partition sequence**. A simple example will show the basic ideas behind this concept. Let us have a simple four atom molecular system where an arbitrary order relation (numbering) has been defined on the set of atoms, and let their  $\{x, y, z\}$  coordinates be

$$\begin{aligned} 1. & \{-17.14, 43.03, 0.50\} \\ 2. & \{-30.35, 44.92, -2.98\} \\ 3. & \{-28.22, 42.82, 8.89\} \\ 4. & \{-21.02, 37.51, -5.19\} \end{aligned} \quad (1)$$

from (1) we see that the following relations hold :

$$x_2 < x_3 < x_4 < x_1, \quad y_4 < y_3 < y_1 < y_2, \quad z_4 < z_2 < z_1 < z_3 \quad (2)$$

from relations (2) we say that, for example, coordinate  $x_3$  **dominates**  $x_2$  and is **dominated** by  $x_4$  and  $x_1$ . Thus the molecular conformation defined by coordinate set (1) can be characterized by the following **dominance partition sequence** ( $DPS$ )

$$\{\{(2)(3)(4)(1)\}_x, \{(4)(3)(1)(2)\}_y, \{(4)(2)(1)(3)\}_z\} \quad (3)$$

where the atom numbers in  $x, y$  and  $z$  are ordered as in (2). From (3) we can partition the ( $CS$ ) of a molecular system into a set of discrete cells such that the conformations in a cell all have the same  $DPS$ , as was discussed in [1,4] cells in  $CS$  have the shape of polyhedral cones with the vertex at the origin.

The formula (3) for  $DPS$ s has the peculiar characteristic that each index is enclosed between parenthesis, this is so because in this way it can be extended to designate sets of contiguous cells. Suppose we have a conformation with  $DPS \{\dots(i)(j)\dots\}_c$ , exchanging the coordinates  $c_i$  and  $c_j$  results in a new conformation with  $DPS \{\dots(j)(i)\dots\}_c$  that lies in an adjacent cell, thus the notation for  $DPS$  in (3) can be extended to

$$\{\dots(i\ j)\dots\}_c \quad (4)$$

where (4) designates all the sequences that can be obtained by a permutation of the consecutive numbers  $i$  and  $j$ .

This can be generalized [4] for any set of consecutive numbers

$$\{\dots(i_1 i_2 \dots i_n)\dots\}_c \quad (5)$$

where (5) designates any sequence obtained by permuting the consecutive numbers  $i_1, i_2 \dots i_n$ . In what follows sequences like (3) will be designated as **simple**  $DPS$ s while (4) and (5) will be **extended**  $DPS$ s.

An inclusion relation among  $DPS$ s can be defined : if  $\mathcal{P}_a$  and  $\mathcal{P}_b$  are two  $DPS$ s and if  $\mathcal{S}_{\mathcal{P}_a}$  and  $\mathcal{S}_{\mathcal{P}_b}$  are the sets of simple sequences that they encode, then

**Definition 1.**  $\mathcal{P}_a \subset \mathcal{P}_b$  if  $\mathcal{S}_{\mathcal{P}_a} \subset \mathcal{S}_{\mathcal{P}_b}$ .

with this definition extended  $DPS$ s amount to more than just sequences: let  $\mathcal{P}$  be any  $DPS$  the

set  $\{\mathcal{S}_P : \forall \mathcal{P}_s \subset \mathcal{P} \Rightarrow \mathcal{P}_s \in \mathcal{S}_P\}$ ,  $\mathcal{S}_P$  is the partially ordered set (**poset**) associated to  $\mathcal{P}$  [1,6,7].

## II. GENERALIZED PARTITION SEQUENCES

DPSs are very useful structures: with a minimum of code they allow to designate huge numbers of cells in *CS*. However, it was shown in [5] that if dominance partition sequences are to quantify the regions enclosing molecular dynamics trajectories they require one further level of codification: parenthesis in (3) and (5) are to be allowed to overlap. For example

$$\{(3 \overset{1}{4} \overset{2}{(8 \overset{1}{9})} 1 \overset{2}{7})\} \quad (6)$$

simultaneously encodes the extended *DPS*s  $\{(3 \ 4 \ 8 \ 9)(1 \ 7)\}$  and  $\{(3 \ 4)(1 \ 7 \ 8 \ 9)\}$ . As it can be seen from (6) pairs of parenthesis must bear an index so we can tell the beginning and the end. In order to distinguish (6) from ordinary *DPS*s the denomination of **generalized dominance partition sequences** (*GDPS*) was proposed in [5].

Coding sets of cells from conformational space as in (6) allows to define a cone in *CS* that wraps the region where the system evolves. For example, from a 2 ns molecular dynamics trajectory of a 58 residue protein structure [9]: the pancreatic trypsin inhibitor [8], using the dominance relations matrix from the figure 1 in [5], we have calculated the minimal *GDPS*-cone enclosing the  $\alpha$ -carbons coordinates of the PTI

$$\begin{aligned} & \{ \{ (29 \overset{1}{48} \overset{2}{49} (\overset{3}{27} (\overset{4}{28} (\overset{5}{30} \overset{1}{31} \overset{2}{52} (\overset{3}{53}) \overset{4}{47} \overset{5}{50}) \overset{6}{32}) (\overset{7}{26}) \overset{8}{21} \overset{9}{51} (\overset{10}{19} \overset{11}{23} \overset{12}{24} (\overset{13}{20} \overset{14}{25} \overset{15}{33}) \overset{16}{46} (\overset{17}{54} \overset{18}{55}) \overset{19}{22}) \\ & \quad 18 \overset{20}{34} (\overset{21}{45}) \overset{22}{17}) (\overset{23}{44} (\overset{24}{5} (\overset{25}{8} (\overset{26}{6} \overset{27}{35}) (\overset{28}{9}) \overset{29}{43}) \overset{30}{16} (\overset{31}{11}) \overset{32}{7}) \overset{33}{36} (\overset{34}{3} \overset{35}{4} \overset{36}{10}) (\overset{37}{37} \overset{38}{42}) (\overset{39}{15} (\overset{40}{12}) \\ & \quad (\overset{41}{41}) \overset{42}{40}) \overset{43}{38} (\overset{44}{14}) \overset{45}{13}) \overset{46}{39}) \}_x, \\ & \{ (15 \overset{1}{16} (\overset{2}{17}) (\overset{3}{14}) (\overset{4}{18}) (\overset{5}{36} (\overset{6}{13}) \overset{7}{37}) (\overset{8}{19} (\overset{9}{34}) \overset{10}{12} \overset{11}{35} \overset{12}{38}) (\overset{13}{11}) \overset{14}{20} \overset{15}{33} (\overset{16}{39}) (\overset{17}{46} (\overset{18}{10}) \overset{19}{32} \overset{20}{40} \overset{21}{47}) \\ & \quad (\overset{22}{21}) (\overset{23}{45}) \overset{24}{44} (\overset{25}{31} (\overset{26}{9} \overset{27}{48}) \overset{28}{41}) (\overset{29}{22}) (\overset{30}{42} \overset{31}{49} \overset{32}{50}) \overset{33}{8} \overset{34}{30} \overset{35}{43} \overset{36}{51}) (\overset{37}{23} (\overset{38}{24}) (\overset{39}{7} \overset{40}{52} (\overset{41}{29} (\overset{42}{4} \overset{43}{53} \overset{44}{54}) \\ & \quad (\overset{45}{26} \overset{46}{27}) \overset{47}{5}) \overset{48}{6} \overset{49}{25} \overset{50}{28} \overset{51}{55}) \overset{52}{3}) \}_y, \\ & \{ (26 \overset{1}{27} (\overset{2}{8} \overset{3}{10} (\overset{4}{7} (\overset{5}{25} (\overset{6}{11} (\overset{7}{13}) \overset{8}{9}) \overset{9}{6} \overset{10}{24} (\overset{11}{12} (\overset{12}{28}) \overset{13}{33}) (\overset{14}{31}) \overset{15}{34} (\overset{16}{15}) (\overset{17}{32} (\overset{18}{29} (\overset{19}{17}) (\overset{20}{14}) \overset{21}{5} \overset{22}{23} \overset{23}{4} \\ & \quad (\overset{24}{22} \overset{25}{35} \overset{26}{36} \overset{27}{40} (\overset{28}{41} \overset{29}{3}) (\overset{30}{30}) \overset{31}{39}) \overset{32}{16} \overset{33}{21}) \overset{34}{43}) \overset{35}{38}) (\overset{36}{18} \overset{37}{19}) \overset{38}{20}) \overset{39}{37}) \overset{40}{42} \overset{41}{44}) (\overset{42}{55}) (\overset{43}{45} \overset{44}{48} (\overset{45}{52}) \overset{46}{51}) \\ & \quad (\overset{47}{46} \overset{48}{47}) \overset{49}{54} (\overset{50}{49} \overset{51}{53}) \overset{52}{5}) \}_z \} \end{aligned} \quad (7)$$

As stressed at the end of section I the importance of expressions like (7) lies in their associated poset: a hierarchical structure, because it allows a hierarchical decomposition of the *CS* region into sets of smaller cones and, more important, it facilitates the sorting of the **graph of cells** [3-5] which is the structure containing all the information about cells in *CS*.

## III. MERGING CELLS FROM ADJACENT NODES IN THE COMPACT GRAPH OF CELLS

The **graph of cells**, or **G** is the fundamental structure of the present approach, it arose from the notion that the relative motions of small sets of atoms in the molecule can be thoroughly sampled in computer simulations. **G** encodes the set of cells in conformational space that can be accessed through the allowed combinations of these movements.

This structure is constructed first by dividing the molecule into four-atom ordered sets whose 3D-structure is that of a **simplex**<sup>1</sup>, next for each simplex the visited cells in its *CS* are determined empirically from computer simulations. Then **G** is built as follows [3-4]

- Each cell from a simplex is a node of **G**
- The edges of the graph are between compatible nodes in adjacent simplexes: two cells from the *CS*s of two simplexes that share a face, or equivalently three atoms, are said to be **compatible** if their *DPS*s restricted to the numbers of the shared atoms are equal [5].
- For every pair of simplexes sharing a face, each cell from the *CS* of one simplex has an edge towards at least one cell from the other simplex, otherwise the structure would be geometrically inconsistent.

**Lemma 1.** *Given a simplex  $\mathcal{S} \in \mathbf{G}$  and a cell  $\xi_1 \in \mathcal{S}$  there is at least one other cell  $\xi_2 \in \mathcal{S}$  such that their respective *DPS*s differ only by a permutation of two numbers.*

Let  $\mathcal{S} = \{n_{s_1}, n_{s_2}, n_{s_3}, n_{s_4}\}$ , assume that we have two cells whose coordinate *c* *DPS*s are  $\{(n_{s_1})(n_{s_2})(n_{s_3})(n_{s_4})\}_c$  and  $\{(n_{s_1})(n_{s_4})(n_{s_2})(n_{s_3})\}_c$  respectively, the remaining coordinates being equal, and that there is not a cell with sequence  $\{(n_{s_1})(n_{s_2})(n_{s_4})(n_{s_3})\}_c$ , this is geometrically impossible because the coordinate  $c_{n_{s_4}}$  cannot pass through  $c_{n_{s_2}}$  without first going through  $c_{n_{s_3}}$ .

There is a class of subgraphs of **G** called **transversals** such that they have a set of nodes consisting of one cell from every simplex and each cell has a edge towards every cell in adjacent simplexes. It was shown in [5] that each cell in a transversal is the projection of one cell from the *CS* of the molecule, thus **G** makes possible the enumeration of accessible cells in the conformational space of a molecule.

**G** can be put in a compact form called **C** by recursively agregating sets of *DPS* in **G** into extended ones that contain them. For instance the cells from the simplex  $\{9, 10, 14, 15\}$  in **G** [4,5]

$$\begin{aligned} & \{ \{ (9)(10)(15)(14) \}_x, \{ (15)(14)(10)(9) \}_y, \{ (9)(10)(14)(15) \}_z \}, \\ & \{ \{ (9)(10)(15)(14) \}_x, \{ (15)(14)(10)(9) \}_y, \{ (9)(10)(15)(14) \}_z \}, \\ & \{ \{ (9)(10)(15)(14) \}_x, \{ (15)(14)(10)(9) \}_y, \{ (10)(15)(9)(14) \}_z \}, \\ & \{ \{ (9)(10)(15)(14) \}_x, \{ (15)(14)(10)(9) \}_y, \{ (10)(9)(15)(14) \}_z \}, \\ & \{ \{ (9)(10)(15)(14) \}_x, \{ (15)(14)(10)(9) \}_y, \{ (10)(9)(14)(15) \}_z \}, \\ & \{ \{ (9)(15)(10)(14) \}_x, \{ (15)(14)(10)(9) \}_y, \{ (9)(10)(14)(15) \}_z \}, \\ & \{ \{ (9)(15)(10)(14) \}_x, \{ (15)(14)(10)(9) \}_y, \{ (9)(10)(15)(14) \}_z \}, \\ & \{ \{ (9)(15)(10)(14) \}_x, \{ (15)(14)(10)(9) \}_y, \{ (10)(15)(9)(14) \}_z \}, \\ & \{ \{ (9)(15)(10)(14) \}_x, \{ (15)(14)(10)(9) \}_y, \{ (10)(9)(15)(14) \}_z \}, \\ & \{ \{ (9)(15)(10)(14) \}_x, \{ (15)(14)(10)(9) \}_y, \{ (10)(9)(14)(15) \}_z \}, \\ & \{ \{ (10)(9)(15)(14) \}_x, \{ (15)(14)(10)(9) \}_y, \{ (9)(10)(14)(15) \}_z \}, \\ & \{ \{ (10)(9)(15)(14) \}_x, \{ (15)(14)(10)(9) \}_y, \{ (9)(10)(15)(14) \}_z \}, \\ & \{ \{ (10)(9)(15)(14) \}_x, \{ (15)(14)(10)(9) \}_y, \{ (10)(15)(9)(14) \}_z \}, \\ & \{ \{ (10)(9)(15)(14) \}_x, \{ (15)(14)(10)(9) \}_y, \{ (10)(9)(15)(14) \}_z \}, \\ & \{ \{ (10)(9)(15)(14) \}_x, \{ (15)(14)(10)(9) \}_y, \{ (10)(9)(14)(15) \}_z \} \} \end{aligned} \quad (8)$$

can be put in a compact form in **C**

$$\{ \{ (9)(10 \ 15)(14) \}_x, \{ (15)(14)(10)(9) \}_y, \{ (10)(9 \ 15)(14) \}_z \},$$

---

<sup>1</sup>An irregular polytope with four vertices.

$$\begin{aligned}
& \{ \{(9\ 10)(15)(14)\}_x, \{(15)(14)(10)(9)\}_y, \{(10)(9\ 15)(14)\}_z \}, \\
& \{ \{(9)(10\ 15)(14)\}_x, \{(15)(14)(10)(9)\}_y, \{(9\ 10)(14\ 15)\}_z \}, \\
& \{ \{(9\ 10)(15)(14)\}_x, \{(15)(14)(10)(9)\}_y, \{(9\ 10)(14\ 15)\}_z \}
\end{aligned} \tag{9}$$

The notion of compatibility can be extended to the cells in  $\mathbf{C}$ : two cells in adjacent simplexes in  $\mathbf{C}$  are said to be **compatible** if the dominance patterns that result from restricting their respective *DPS* sequences to the atom numbers from the shared face are equal.

#### IV. THE FRAGMENTATION OF THE CONE

The cone described in section II encloses the volume in *CS* that the system can access, this cone has been built by looking empirically in computer simulations at the range of each atom coordinate independently [5], so part of the volume is wasted as the construction of the cone does not take into account how the *x*, *y* and *z* coordinates are correlated. The graph  $\mathbf{C}$  can supply the missing information allowing us to create cones that wrap more closely to the cells in it.

We propose here an approach to the fragmentation of the cone that is based on the following set of heuristic rules:

1. For a simplex  $\mathcal{S} \in \mathbf{C}$  we scan the cell sequences for each coordinate independently: if for the coordinate *c* the partition sequences of two cells  $\xi_1, \xi_2 \in \mathcal{S}$  are such that  $\mathcal{P}_{1,c} \subset \mathcal{P}_{2,c}$ , then we set  $\mathcal{P}_{1,c} := \mathcal{P}_{2,c}$ .
2. Let  $\mathcal{S} = \{n_{s_1}, n_{s_2}, n_{s_3}, n_{s_4}\}$  be a simplex in  $\mathbf{C}$  and be  $\{\mathcal{P}_{x,c} : 1 \leq x \leq N_\xi\}$  its set of cell sequences in the *c* coordinate. For 4 numbers there is a set of 24 possible simple *DPS*s, if every sequence from this set is contained in at least one  $\mathcal{P}_{x,c}$ , then  $\forall x : 1 \leq x \leq N_\xi$  we set  $\mathcal{P}_{x,c} := \{(n_{s_1} \ n_{s_2} \ n_{s_3} \ n_{s_4})\}_c$ .
3. After performing the two previous steps for *x*, *y* and *z* redundant sequences are removed.

As an example the set of cells for the simplex  $\mathcal{S} = \{3, 4, 5, 8\}$  in  $\mathbf{C}$  are

$$\begin{aligned}
& \{ \{(5\ 8)(4)(3)\}_x, \{(8)(4)(3\ 5)\}_y, \{(8)(4\ 5)(3)\}_z \}, \\
& \{ \{(5\ 8)(4)(3)\}_x, \{(8)(4)(3\ 5)\}_y, \{(8)(5)(3\ 4)\}_z \}, \\
& \{ \{(5)(8)(3\ 4)\}_x, \{(8)(4)(3\ 5)\}_y, \{(8)(4\ 5)(3)\}_z \}, \\
& \{ \{(5\ 8)(3\ 4)\}_x, \{(8)(4\ 5)(3)\}_y, \{(8)(3\ 4\ 5)\}_z \}
\end{aligned} \tag{10}$$

applying the above transformation gives

$$\begin{aligned}
& \{ \{(5\ 8)(3\ 4)\}_x, \{(8)(4)(3\ 5)\}_y, \{(8)(3\ 4\ 5)\}_z \}, \\
& \{ \{(5\ 8)(3\ 4)\}_x, \{(8)(4\ 5)(3)\}_y, \{(8)(3\ 4\ 5)\}_z \}
\end{aligned} \tag{11}$$

(11) is a cone that contains (10), in *GDPS* notation

$$\{ \overset{1}{(5)} \overset{12}{(8)} \overset{2}{(3\ 4)} \}_x, \{ \overset{1}{(8)} \overset{1\ 2}{(4\ 5)} \overset{3}{(3)} \}_y, \{ \overset{1}{(8)} \overset{1\ 2}{(3\ 4\ 5)} \}_z \} \tag{12}$$

This operation is performed in order to determine the minimal set of ordinary cones that contain the cells from a given simplex.

One aim of this note is to show that the set of rules described above give meaningful results, a detailed study of the decomposition of (9) will be the subject of a further communication. Here, we show the above procedure at work on a simpler case, that of the first 13  $\alpha$ -carbons from our example molecule, they have the *GDPS* sequence

$$\begin{aligned}
& \{(5 \overset{1}{(8 \overset{2}{(6 \overset{3}{(9 \overset{4}{11 \overset{1}{7}})})})}) (3 \overset{4}{4} \overset{6}{10} \overset{5}{15} \overset{7}{(12)} \overset{6}{(14)} \overset{7}{13})\}_x, \\
& \{(15 \overset{1}{(14 \overset{1}{(13 \overset{2}{(11 \overset{2}{12 \overset{3}{(10 \overset{3}{(9 \overset{4}{(8 \overset{5}{(7 \overset{6}{(4 \overset{7}{(5 \overset{8}{6 \overset{9}{3}})})})})})})})})})})})\}_y, \\
& \{(7 \overset{1}{8} \overset{2}{10} \overset{3}{(6 \overset{3}{9} \overset{4}{11} \overset{1}{13} \overset{2}{12})} \overset{3}{(15)} \overset{4}{14} \overset{4}{3} \overset{4}{4} \overset{4}{5})\}_z
\end{aligned} \tag{13}$$

We can see a total of 7 sequences of permuting numbers in  $x$ , 10 in  $y$  and 4 in  $z$ , our interest here is to see the realizable combinations of them obtained by merging compatible transformed sequences from different simplexes.

A first example is from the  $x_4$  sequence that has 6 numbers in it (3 4 7 9 10 11), this means that the basic information about it, that is: how it connects with sequences in other dimensions, is contained in 15 simplexes, whose transformed *DPSs* are

$$\begin{aligned}
& \{(3 \ 4 \ 7 \ 9)\}_x, \{(9)(4 \ 7)(3)\}_y, \{(7 \ 9)(3 \ 4)\}_z \\
& \{(3 \ 4 \ 7 \ 10)\}_x, \{(10)(4 \ 7)(3)\}_y, \{(7 \ 10)(3 \ 4)\}_z \\
& \{(3 \ 4 \ 7 \ 11)\}_x, \{(11)(4 \ 7)(3)\}_y, \{(7 \ 11)(3 \ 4)\}_z \\
& \{(3 \ 4 \ 9 \ 10)\}_x, \{(10)(9)(4)(3)\}_y, \{(9 \ 10)(3 \ 4)\}_z \\
& \{(3 \ 4 \ 9 \ 11)\}_x, \{(11)(9)(4)(3)\}_y, \{(9 \ 11)(3 \ 4)\}_z \\
& \{(3 \ 4 \ 10 \ 11)\}_x, \{(11)(10)(4)(3)\}_y, \{(10 \ 11)(3 \ 4)\}_z \\
& \{(3 \ 7 \ 9 \ 10)\}_x, \{(10)(9)(7)(3)\}_y, \{(7 \ 9 \ 10)(3)\}_z \\
& \{(3 \ 7 \ 9 \ 11)\}_x, \{(11)(9)(7)(3)\}_y, \{(7 \ 9 \ 11)(3)\}_z \\
& \{(3 \ 7 \ 10 \ 11)\}_x, \{(11)(10)(7)(3)\}_y, \{(7 \ 10 \ 11)(3)\}_z \\
& \{(3 \ 9 \ 10 \ 11)\}_x, \{(11)(10)(9)(3)\}_y, \{(9 \ 10 \ 11)(3)\}_z \\
& \{(4 \ 7 \ 9 \ 10)\}_x, \{(10)(9)(4 \ 7)\}_y, \{(7 \ 9 \ 10)(4)\}_z \\
& \{(4 \ 7 \ 9 \ 11)\}_x, \{(11)(9)(4 \ 7)\}_y, \{(7 \ 9 \ 11)(4)\}_z \\
& \{(4 \ 7 \ 10 \ 11)\}_x, \{(11)(10)(4 \ 7)\}_y, \{(7 \ 10 \ 11)(4)\}_z \\
& \{(4 \ 9 \ 10 \ 11)\}_x, \{(11)(10)(9)(4)\}_y, \{(9 \ 10 \ 11)(4)\}_z \\
& \{(7 \ 9 \ 10 \ 11)\}_x, \{(11)(10)(9)(7)\}_y, \{(7 \ 9 \ 10 \ 11)\}_z
\end{aligned} \tag{14}$$

obviously the sequences from these 15 cones are all compatible with those from adjacent simplexes, so they can all be merged to give the result

$$\{(3 \ 4 \ 7 \ 9 \ 10 \ 11)\}_x, \{(11)(10)(9)(4 \ 7)(3)\}_y, \{(7 \ 9 \ 10 \ 11)(3 \ 4)\}_z \tag{15}$$

in which  $x_4$  appears to be connected with sequences  $z_1$  and  $z_4$ , or at least with fragments of them. Contrasting with (15) the sequence  $z_1$ , of length 8, after merging up the *DPSs* from 70 simplexes we end with a total of 7 cones

$$\begin{aligned}
& \{(8 \ 9)(6 \ 7 \ 11)(10)(12)(13)\}_x, \{(13)(11 \ 12)(7)(10)(9)(8)(7)(6)\}_y, \{(6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12 \ 13)\}_z \\
& \{(8 \ 9 \ 11)(6 \ 7)(10)(12)(13)\}_x, \{(13)(11 \ 12)(7)(10)(9)(8)(7)(6)\}_y, \{(6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12 \ 13)\}_z \\
& \{(8)(6 \ 7 \ 9 \ 11)(10)(12)(13)\}_x, \{(13)(11 \ 12)(7)(10)(9)(8)(7)(6)\}_y, \{(6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12 \ 13)\}_z \\
& \{(8 \ 11)(6 \ 7 \ 9)(10)(12)(13)\}_x, \{(13)(11 \ 12)(7)(10)(9)(8)(7)(6)\}_y, \{(6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12 \ 13)\}_z \\
& \{(6 \ 8 \ 11)(7 \ 9 \ 10)(12)(13)\}_x, \{(13)(11 \ 12)(7)(10)(9)(8)(7)(6)\}_y, \{(6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12 \ 13)\}_z \\
& \{(6 \ 8)(7 \ 9 \ 10 \ 11)(12)(13)\}_x, \{(13)(11 \ 12)(7)(10)(9)(8)(7)(6)\}_y, \{(6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12 \ 13)\}_z \\
& \{(6 \ 8 \ 9 \ 11)(7 \ 10)(12)(13)\}_x, \{(13)(11 \ 12)(7)(10)(9)(8)(7)(6)\}_y, \{(6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12 \ 13)\}_z
\end{aligned}$$

$z_1$  appears to be connected with  $x_2$ ,  $x_3$  and  $x_4$ , but there is also 4 cones that combine with fragments from the main sequences. This reveals details of how 3D structures are organized inside the main cone: only some of the *DPSs* in one dimension connect with those from another dimension. The exception in this example are the segments from the  $y$  dimension: there is so few structure in them that they appear to combine with any other segment from  $x$  and  $y$ .

The remaining segments  $x_1$ ,  $x_5$  and  $z_4$  give

$$\{\{(5\ 6\ 8\ 9)\}_x, \{(9)(8)(5\ 6)\}_y, \{(6\ 8\ 9)(5)\}_z\} \quad (x_1)$$

$$\{\{(3\ 4\ 10\ 12\ 15)\}_x, \{(15)(12)(10)(4)(3)\}_y, \{(10\ 12\ 15)(3\ 4)\}_z\} \quad (x_5)$$

$$\begin{aligned} &\{\{(5)(3\ 4\ 15)(14)\}_x, \{(15)(14)(4)(3\ 5)\}_y, \{(3\ 4\ 5\ 14\ 15)\}_z\} \\ &\{\{(5)(3\ 4\ 15)(14)\}_x, \{(15)(14)(4\ 5)(3)\}_y, \{(3\ 4\ 5\ 14\ 15)\}_z\} \end{aligned} \quad (z_4)$$

## VI. CONCLUSION

The *GDPS* (7) is a global approximation that sets the bounds of the region from *CS* in which the system evolves. These bounds are set independently for the  $x$ ,  $y$  and  $z$  coordinates of the molecule, and consequently they are not correlated: this means that much of the volume enclosed by (7) does not correspond to realizable 3D-structures.

On the other hand the graph of cells  $\mathbf{G}$  allows to exactly enumerate the set of visited cells in conformational space, but this possibility is probably algorithmically hopeless.

The approach presented in this paper consists in using the information contained in  $\mathbf{G}$  to derive bounds from (7) that are correlated in  $x$ ,  $y$  and  $z$ , and to progressively narrow these bounds around interesting regions.

The importance of a structure like (7), and the *DPSs* that are derived from them, is not only the precision they can attain in delimiting conformational space, but the fact that they possess a **graphical structure** and connected graphs always have a metric: we can measure the distance between two nodes as the length of the minimal path that joins them. This opens the very real possibility of measuring and enumerating distances between points and conformations, so that a combinatorial hamiltonian can be built on these structures. And this should be the next phase of the present work.

## REFERENCES

1. J. Gabarro-Arpa, "A central partition of molecular conformational space. I. Basic structures", *Comp. Biol. and Chem.* **27**, 153-159 (2003).
2. J. Gabarro-Arpa, "A central partition of molecular conformational space. II. Embedding 3D-structures", *Proceedings of the 26th Annual International Conference of the IEEE EMBS*, San Francisco, 3007-3010 (2004).
3. J. Gabarro-Arpa, "Combinatorial determination of the volume spanned by a molecular system in conformational space", *Lecture Series on Computer and Computational Sciences* **4**, 1778-1781 (2005).
4. J. Gabarro-Arpa, "A central partition of molecular conformational space. III. Combinatorial determination of the volume spanned by a molecular system in conformational space", *Journal of Mathematical Chemistry* DOI 10.1007 s10910-006-9079-8.url (2006).
5. J. Gabarro-Arpa, "A Central Partition of Molecular Conformational Space. IV. Extracting information from the graph of cells", arXiv: physics/061108v2, submitted for publication (2007).
6. S. Fomin and N. Reading, "Root systems and generalized associahedra", math.CO/0505518 (2005).
7. A. Bjorner, M. las Vergnas, B. Sturmfels, N. White, "Oriented Matroids", Cambridge, UK, Cambridge University Press, sect. **2** (1993).
8. M. Marquart, J. Walter, J. Deisenhofer, W. Bode, R. Huber, "The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors", *Acta Crystallogr. Sect. B*, **39**, 480-490 (1983).
9. J. Gabarro-Arpa, R. Revilla, "Clustering of a molecular dynamics trajectory with a Hamming distance", *Comp. and Chem.* , **24**, 693-698 (2000).